

STAT100 Problem Set 4: Examining Relationships

You need to submit a Word document or PDF for this assignment. Make sure you do the following:

1. Upload only one file/document in ELMS for Problem Set 4.
2. Include your name in the document in the upper left-hand corner. Under your name, write STAT 100 and your section number. Write Problem Set 4 centered on the page.
3. Number and letter your answers to the questions accordingly.
4. Carefully read all problems and follow all instructions.
5. Upload the assignment in ELMS before the deadline of Sunday 3/6 at 11:59 PM otherwise it is considered late. Make sure you save your document on your computer or email it to yourself so that you keep an electronic copy.

STAT100 Problem Sets need to be completed by students in RStudio. Students should refer to the Tutorial for Problem Set 4 as they are working on this problem set. All Tutorials for Problem Sets can be found in the STAT100 ELMS course under Modules.

For this problem set you need to use the *Restaurant_Tips.RData*. The *Restaurant_Tips* dataset includes data collected from a restaurant during a five-day period. Below is a description of each variable:

- *Bill* – total amount of bill in dollars, including tax, but not including tip
- *Tip* – total amount of tip left by the customers at the table
- *Credit* – indicates whether the customers used a credit/debit card
- *Guests* – total number of guests at the table
- *Day* – indicates the day of the week
- *Server* – indicates the name of the server
- *PctTip* – the tip as a percent of the total amount of the bill

Problem Set 4 has three questions worth 25 points. Read each question carefully and follow all instructions. **Please follow these instructions for providing your responses:**

- For #1.d., 2.d., and 3.d., you need to provide R code and RStudio output.
- For all other questions on Problem Set 4, you should type your responses directly in the document you submit for Problem Set 4.

In this Problem Set there are three questions related to examining relationships between two variables. For each question you need to first identify the explanatory variable, the response variable, and the “role-type classification”. Next, using RStudio you will generate the appropriate data displays and/or numeric values to examine the potential relationship between the two variables. Lastly you will interpret the data displays and numeric values.

1. (8 points) **Open the *Restaurant_Tips.RData* file in RStudio.** You can use this data set to examine if the tip as a percent of the bill (the *Pct_Tip* variable) can be predicted by the server (the *Server* variable). Use the *Restaurant_Tips.RData* file to examine this potential relationship.
 - a. Identify the explanatory variable.
 - b. Identify the response variable.
 - c. Identify the specific role-type classification (use the shorthand notation for role-type classification).
 - d. In RStudio, generate the appropriate data displays and/or numeric values to investigate this relationship. **See the Tutorial for Problem Set 4 to determine which data displays and/or numeric values are needed for each role-type classification.** In the document you upload for this assignment, for 1.d. include:
 - i. R code and any RStudio output to examine this relationship, including comments (**IMPORTANT INSTRUCTIONS: you MUST include your name in a comment line of the R code for #1.d.**)
 - ii. Images of any data displays generated in RStudio to examine this relationship. **IMPORTANT INSTRUCTIONS: the titles of any data displays (side-by-side boxplots or scatterplots) that you generate in RStudio MUST include your name.**
 - e. During the five-day period, which server had the highest median tip percentage? In your answer, provide the value of the median tip percentage for that server.
 - f. During the five-day period, which server had the lowest median tip percentage? In your answer, provide the value of the median tip percentage for that server.
 - g. Interpret the data displays and/or numeric values that you generated in 1.d. to compare the distribution of the response variable for the different values of the explanatory variable. Based on your interpretation, would you say that there is a relationship between the server and the tip as a percent of the bill? In other words, would you conclude that the different servers seem to receive higher (or lower) tip percentages?

2. (7 points) **Open the *Restaurant_Tips.RData* file in RStudio.** You can use this data set to examine if a server is more likely to work on certain days. In other words, you can examine if the *Server* variable can be predicted by the *Day* variable. Use the *Restaurant_Tips.RData* file to examine this potential relationship.
- Identify the explanatory variable.
 - Identify the response variable.
 - Identify the specific role-type classification (use the shorthand notation for role-type classification).
 - In RStudio, generate the appropriate data displays and/or numeric values to investigate this relationship. **See the Tutorial for Problem Set 4 to determine which data displays and/or numeric values are needed for each role-type classification.** In the document you upload for this assignment, for 2.d. include:
 - R code and any RStudio output to examine this relationship, including comments (**IMPORTANT INSTRUCTIONS: you MUST include your name in a comment line of the R code for #2.d.**)
 - Images of any data displays generated in RStudio to examine this relationship. **IMPORTANT INSTRUCTIONS: the titles of any data displays (side-by-side boxplots or scatterplots) that you generate in RStudio MUST include your name.**
 - During the five-day period, what percentage of tables on Friday were served by each server? Enter the percentages for each server as a percent rounded to two decimal places (e.g., 27.34%).
 - During the five-day period, were there any days that one or more of the servers did not serve any tables? Provide the day and the name or names of the servers who did not serve any tables on that day.
 - Interpret the data displays and/or numeric values that you generated in 2.d. to compare the distribution of the response variable for the different values of the explanatory variable. Based on your interpretation, would you say that there is a relationship between the day of the week and the server? In other words, would you conclude that the different servers are more (or less) likely to serve tables on different days?

3. (10 points) **Open the *Restaurant_Tips.RData* file in RStudio.** You can use this data set to examine if tables tend to leave higher tip percentages for different amounts of the total bill for the table. In other words, you can examine if the *Bill* variable can predict the *Pct_Tip* variable. Use the *Restaurant_Tips.RData* file to examine this potential relationship.
- Identify the explanatory variable.
 - Identify the response variable.
 - Identify the specific role-type classification (use the shorthand notation for role-type classification).
 - In RStudio, generate the appropriate data displays and/or numeric values to investigate this relationship. **See the Tutorial for Problem Set 4 to determine which data displays and/or numeric values are needed for each role-type classification.** In the document you upload for this assignment, for 3.d. include:
 - R code and any RStudio output to examine this relationship, including comments (**IMPORTANT INSTRUCTIONS: you MUST include your name in a comment line of the R code for #3.d.**)
 - Images of any data displays generated in RStudio to examine this relationship. **IMPORTANT INSTRUCTIONS: the titles of any data displays (side-by-side boxplots or scatterplots) that you generate in RStudio MUST include your name.**
 - Provide the value of the correlation coefficient (r) for this potential relationship.
 - What does the value of the correlation coefficient indicate about the strength and direction of the linear relationship between the bill amount and tip percentage?
 - Provide the least-squares linear regression equation for this relationship. **See the Tutorial for Problem Set 4 for guidance on how to write the regression equation.**
 - Use the regression equation to predict the tip percentage for a table with a total bill amount of \$50.00. Show all work and provide your answer as a percent rounded to two decimal places (e.g., 37.24%).
 - Interpret the data displays and/or numeric values that you generated for 3.d. to compare the distribution of the response variable for the different values of the explanatory variable. Based on your interpretation, would you say that there is a relationship between the bill amount and the tip percentage?